

# Macroscopic time evolution and MaxEnt inference for closed systems with Hamiltonian dynamics

Domagoj Kuić,<sup>\*</sup> Paško Županović,<sup>†</sup> and Davor Juretić<sup>‡</sup>

*University of Split, Faculty of Science, N. Tesle 12, 21000 Split, Croatia*

## Abstract

MaxEnt inference algorithm and information theory are relevant for the time evolution of macroscopic systems considered as problem of incomplete information. Two different MaxEnt approaches are introduced in this work, both applied to prediction of time evolution for closed Hamiltonian systems. The first one is based on Liouville equation for the conditional probability distribution, introduced as a strict microscopic constraint on time evolution in phase space. The conditional probability distribution is defined for the set of microstates associated with the set of phase space paths determined by solutions of Hamilton's equations. The MaxEnt inference algorithm with Shannon's concept of the conditional information entropy is then applied to prediction, consistently with this strict microscopic constraint on time evolution in phase space. The second approach is based on the same concepts, with a difference that Liouville equation for the conditional probability distribution is introduced as a macroscopic constraint given by a phase space average. We consider the incomplete nature of our information about microscopic dynamics in a rational way that is consistent with Jaynes' formulation of predictive statistical mechanics, and the concept of macroscopic reproducibility for time dependent processes. Maximization of the conditional information entropy subject to this macroscopic constraint leads to a loss of correlation between the initial phase space paths and final microstates. Information entropy is the theoretic upper bound on the conditional information entropy, with the upper bound attained only in case of the complete loss of correlation. In this alternative approach to prediction of macroscopic time evolution, maximization of the conditional information entropy is equivalent to the loss of statistical correlation, and leads to corresponding loss of information. In accordance with the original idea of Jaynes, irreversibility appears as a consequence of gradual loss of information about possible microstates of the system.

PACS numbers: 02.50.Tt, 05.20.Gg, 05.70.Ln, 45.20.Jj

Keywords: MaxEnt algorithm, information theory, statistical mechanics, Hamiltonian dynamics

---

<sup>\*</sup>Electronic address: [dkuic@pmfst.hr](mailto:dkuic@pmfst.hr)

<sup>†</sup>Electronic address: [pasko@pmfst.hr](mailto:pasko@pmfst.hr)

<sup>‡</sup>Electronic address: [juretic@pmfst.hr](mailto:juretic@pmfst.hr)

## I. INTRODUCTION

Maximum-entropy formalism, or alternatively MaxEnt algorithm, was formulated by E. T. Jaynes in his influential papers [1, 2] intended for applications in statistical mechanics. In Jaynes' approach a full development of the results of equilibrium statistical mechanics and formalism of Gibbs [3] was possible as a form of statistical inference based on Shannon's concept of information-theory entropy as a measure of information [4]. In the language of Jaynes, it is the correct measure of the "amount of uncertainty" in the probability distribution [5]. Maximization of information-theory entropy subject to certain constraints is a central concept in Jaynes' approach, and provides the least biased probability estimates subject to the available information. It is important that Jaynes sees Gibbs' formalism as essential tool for statistical inference in different problems with insufficient information. This includes equilibrium statistical mechanics [1] and the formulation of a theory of irreversibility [2], that Jaynes tries to accomplish in his later works [5–9].

Predictions and calculations for different irreversible processes usually involve three distinct stages [7]: (1) Setting up an "ensemble", i.e., choosing an initial density matrix, or in our case an  $N$ -particle distribution function, which is to describe our initial knowledge about the system of interest; (2) Solving the dynamical problem; i.e., applying the microscopic equations of motion to obtain the time evolution of the system; (3) Extracting the final physical predictions from the time developed ensemble. As fully recognized by Jaynes, the stage (1) and the availability of its general solution simplifies the complicated stage (2). The problem includes also an equally important stage (0) consisting of some kind of measurement or observation defining both the system and the problem [10]. In direct mathematical attempts that lead to a theory of irreversibility, the Liouville theorem with the conservation of phase space volume inherent to Hamiltonian dynamics, is represented as one of the main difficulties. Relation of the Liouville equation and irreversible macroscopic behavior is one of the central problems in statistical mechanics. For this reason it is reduced to an irreversible equation termed Boltzmann equation, rate equation or master equation. Far from creating difficulties, Jaynes considers the Liouville equation and the related constancy in time of Gibbs' entropy as precisely the dynamical property needed for solution of this problem, considering it to be more of conceptual than mathematical nature [6].

Mathematical clarity of this viewpoint has its basis in a limit theorem noted by Shannon [4], an application of the fundamental asymptotic equipartition theorem of information theory. This theorem relates the Boltzmann's original formula for entropy of a macrostate and the Gibbs expression for entropy in the limit of a large number of particles [6, 7, 9]. Mathematical connection with the

Boltzmann's interpretation of entropy as the logarithm of the number of ways (or microstates) by which a macroscopic state can be realized,  $S = k \log W$ , introduces then a simple physical interpretation to the Gibbs' formalism, and its generalizations in the maximum-entropy formalism. Maximization of the information entropy subject to constraints then predicts the macroscopic behavior that can happen in the greatest number of ways compatible with the available information. In application to time dependent processes, this is referred to by Jaynes as the maximum caliber principle [8, 9]. Jaynes clearly stated that this does not represent a physical theory that explains the behavior of different systems by deductive reasoning from the first principles, but a form of statistical inference that makes predictions of observable phenomena from incomplete information [8]. For this reason predictive statistical mechanics can not claim deductive certainty for its predictions. This does not mean that it ignores the laws of microphysics; it certainly uses everything known about the structure of microstates and any data on macroscopic quantities, without making any extra physical assumptions beyond what is given by available information. It is important to note that sharp, definite predictions of macroscopic behavior are possible only when it is characteristic of each of the overwhelming majority of microstates compatible with data. For the same reason, this is just the behavior that is reproduced experimentally under those constraints; this is known essentially as the principle of macroscopic uniformity [1, 2], or reproducibility [9]. In somewhat different context this property is recognized as the concept of macroscopic determinism, whose precise definition involves some sort of thermodynamic limit [11].

In Jaynes' view, the dynamical invariance of the Gibbs' entropy gives a simple proof of the second law, which is then a special case of a general requirement for any macroscopic process to be experimentally reproducible [6]. In the simple demonstration based on the Liouville theorem, this makes possible for Jaynes to generalize the second law beyond the restrictions of initial and final equilibrium states, by considering it a special case of a general restriction on the direction of any reproducible process [6, 12]. The real reason behind the second law, since phase space volume is conserved in the dynamical evolution, is a fundamental requirement on any reproducible process that the phase space volume  $W'$ , compatible with the final state, can not be less than the phase space volume  $W_0$  which describes our ability to reproduce the initial state [6]. The arguments used in this demonstration imply also the question how to determine which nonequilibrium or equilibrium states can be reached from others, and this is not possible without information about dynamics, constants of motion, constraints, etc. The second law predicts only that a change of macroscopic state will go in the general direction of greater final entropy [9]. Better predictions are possible only by introducing more information. Macrostates of higher entropy can be realized

in overwhelmingly more ways, and this is the reason for high reliability of the Gibbs equilibrium predictions [9]. In this context, Jaynes also speculated that accidental success in reversal of an irreversible process is exponentially improbable [12].

Jaynes' interpretation of irreversibility and the second law reflects the point of view of the actual experimenter. Zurek [13] has introduced algorithmic randomness as the measure of the complexity of the microscopic state. He has prescribed entropy not only to the ensemble but also to the microscopic state. This prescription makes the principal distinction between his and Jaynes' approach. The basic laws of computation reflected in this interpretation allow Zurek to formulate thermodynamics from the point of view of Maxwell demon-type entities that can acquire information through measurements and process it in a manner analogous to Turing machines. According to Jaynes, the detailed description of microscopic development of the system can not be extracted from the data about macroscopic development, and therefore it is not a subject of his approach. Increase of entropy is related to gradual decrease of information about possible microstates of the system compatible with data. The notion that the second law is a law of information dynamics, operating at the level of "information bookkeeping", has been considered recently by Duncan and Semura [14, 15]. In this line of thinking, the dynamics of information is considered to be coupled, but fundamentally independent of energy dynamics.

MaxEnt algorithm and its methods represent a way of assigning probability distributions with the largest uncertainty and extent compatible with the available information, and for the same reasons, least biased with respect to unavailable information. Inferences drawn in this way depend only on our state of knowledge [1, 2]. In this work, two different applications of MaxEnt algorithm to macroscopic closed systems with Hamiltonian dynamics, and their time evolution, are examined in detail along with their consequences. The concepts of *phase space paths* with the *path probability distribution* and associated *conditional probability distribution* are defined. The respective *path information entropy* and *conditional information entropy* are introduced in correspondence with definitions in Shannon's information theory [4]. In the first approach, Liouville equation for the conditional probability distribution is introduced as a strict *microscopic constraint* on the time evolution in phase space, which is then completely determined by this constraint and initial values. Maximization of the conditional information entropy, subject to this constraint, predicts the macroscopic behavior that can happen in the greatest number of ways consistent with the information about microscopic dynamics. If probabilities are considered in the objective sense as a property of the system and not of our state of knowledge, full justification of this approach is possible only if our knowledge of the microscopic dynamics is complete. In a similar line of

reasoning Grandy [16] has developed a detailed model of time dependent probabilities for macroscopic systems within MaxEnt formalism and applied it to typical processes in nonequilibrium thermodynamics and hydrodynamics [17, 18]. In a context of the interplay between macroscopic constraints on the system and its microscopic dynamics, it is interesting to note that MaxEnt has been also studied as a method of approximately solving partial differential equations governing the time evolution of probability distribution functions. For more complete further reference, we only mention here that this method, among other examples, has been applied to the Liouville–von Neumann equation [19], the family of dynamical systems with divergenceless phase space flows including Hamiltonian systems [20], the generalized Liouville equation and continuity equations [21]. Universality of this approach has been established for the general class of evolution equations that conform to the essential requirements of linearity and preservation of normalization [22]. This method has been also considered for classical evolution equations with source terms within a framework where normalization is not preserved [23].

The described first approach allows us to define concepts that are basis for our second approach. The difference is that Liouville equation for the conditional probability distribution is now introduced as a *macroscopic constraint*. This constraint on time evolution of the phase space probability density functions is now taken only on average, and it is given by the integral over accessible phase space. It is similar in this respect to constraints given by the data on macroscopic quantities. In Jaynes’ predictive statistical mechanics more objectivity is ascribed to experimentally measured quantities than to probability distributions. The subjective aspect that becomes important here is that probabilities are assigned because of incomplete knowledge, i.e., partial information, and therefore represent our state of knowledge about the system. If information about dynamics is not sufficient to determine the time evolution, an average is taken over all cases possible on the basis of partial information. It is observed how elements of irreversible macroscopic behavior in closed systems with Hamiltonian dynamics are then a consequence of gradual loss of information about possible microstates of the system. This idea has been developed by Jaynes in the density-matrix formalism [2]. In the approach which is developed here, we show that irreversible macroscopic behavior and Jaynes’ interpretation based on reproducibility and information loss, has a clear mathematical description in the concepts of conditional information entropy, and its relation with the information entropy, i.e., in concepts of Shannon’s information theory [4]. At the end of the work, the subjective and objective aspects of both approaches are indicated, and relations with entropy production are discussed.

## II. HAMILTONIAN DYNAMICS AND PHASE SPACE PATHS

The dynamical state of a Hamiltonian system with  $s$  degrees of freedom is described by the coordinates  $q_1, q_2, \dots, q_s$  and the momenta  $p_1, p_2, \dots, p_s$ . At any time  $t$  it is represented by a point in the  $2s$ -dimensional Euclidean space  $\Gamma$  here for our purposes called the phase space of the system. The notation  $(q, p)$  is introduced for the set of  $2s$  coordinates and momenta. The time dependence of  $2s$  dynamical variables  $(q, p)$  is determined by Hamilton's equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad 1 \leq i \leq s, \quad (1)$$

where  $H \equiv H(q, p)$  is the Hamiltonian function of the system. Given the values  $(q_0, p_0)$  at some time  $t_0$ , the solution of Hamilton's equations (1) uniquely determines the values of dynamical variables  $(q, p)$  at any other time  $t$ ,

$$q_i = q_i(t; q_0, p_0), \quad p_i = p_i(t; q_0, p_0), \quad 1 \leq i \leq s. \quad (2)$$

Any point  $(q, p)$  in the phase space  $\Gamma$  describes a curve called a *phase space path*, given by the uniquely determined solution of (1). At any time  $t$  through each point of  $\Gamma$  passes only one path, and this is denoted by the index in  $(q, p)_\omega$ , where  $\omega \in \Omega(\Gamma)$ . The set  $\Omega(\Gamma)$  is the set of all paths in  $\Gamma$ . The velocity  $v$  of the point  $(q, p)$  in the phase space  $\Gamma$  at time  $t$  is given by

$$v \equiv |\mathbf{v}| = \sqrt{\sum_{i=1}^s \left(\frac{dq_i}{dt}\right)^2 + \sum_{i=1}^s \left(\frac{dp_i}{dt}\right)^2} = \sqrt{\sum_{i=1}^s \left(\frac{\partial H}{\partial p_i}\right)^2 + \sum_{i=1}^s \left(\frac{\partial H}{\partial q_i}\right)^2}. \quad (3)$$

The velocity vector  $\mathbf{v}((q, p)_\omega, t)$  is tangential at the point  $(q, p)_\omega \in \Gamma$  to the phase space path  $\omega$  passing through it at time  $t$ . For the systems considered here the Hamiltonian function  $H(q, p)$  does not depend on time and the velocity field  $\mathbf{v}(q, p, t)$  of all points in the phase space  $\Gamma$  is stationary, i.e.,  $\mathbf{v}(q, p, t) = \mathbf{v}(q, p)$ .

Let  $M_0$  be any measurable (in the sense of Lebesgue) set of points in the phase space  $\Gamma$ . In the Hamiltonian motion the set  $M_0$  is transformed into another set  $M_t$  during an interval of time  $t$ . Liouville's theorem asserts that the measure of the set  $M_t$  for any  $t$  coincides with the measure of the set  $M_0$  [24, pp. 15–16]. This theorem proves that the measure in the phase space  $\Gamma$ ,

$$\mu(M_t) = \int_{M_t} dq_1 \dots dq_s dp_1 \dots dp_s = \int_{M_t} d\Gamma, \quad (4)$$

is invariant under Hamiltonian motion. In the notation used in (4), volume element  $dq_1 \dots dq_s dp_1 \dots dp_s$  of the phase space  $\Gamma$  is denoted by  $d\Gamma$ . One immediate corollary [24, pp. 18–19] of Liouville's theorem is that, if  $M_0$  is a Lebesgue measurable set of points of the phase

space  $\Gamma$ , of finite measure, and  $f(q, p)$  a phase function Lebesgue integrable over  $\Gamma$ , then

$$\int_{M_t} f(q, p) d\Gamma = \int_{M_0} f(q(t; q_0, p_0), p(t; q_0, p_0)) d\Gamma_0. \quad (5)$$

Equation (5) is obtained by changing the variables in the integral and introducing new variables  $(q_0, p_0)$ , related to the variables  $(q, p)$  by transformation of the space  $\Gamma$  into itself in Hamiltonian motion, given by (2). If, in particular, the set  $M$  is invariant to the Hamiltonian motion, then using this corollary, it is easy to show how an integral of a phase function  $f(q, p)$  over the invariant set  $M$  is transformed into an integration over the set  $\Omega(M)$  of all paths in  $M$ . This procedure is now developed in the rest of this section. It is used in the definition of probability distributions in Sect. III.

At any time  $t$  through each point  $(q, p)_\omega \in \Gamma$  passes only one path  $\omega \in \Omega(\Gamma)$ , that also passes through the point  $(q_0, p_0)_\omega \in \Gamma$  given by the inverse of (2). The infinitesimal volume element  $d\Gamma_0$  around the point  $(q_0, p_0)_\omega$  can be written as  $d\Gamma_0 = ds_{0\omega} dS_{0\omega}$ . Here,  $ds_{0\omega}$  is the infinitesimal distance along the path  $\omega$ . The infinitesimal element  $dS_{0\omega}$  of the surface  $S_0(M)$  intersects the path  $\omega$  perpendicularly at the point  $(q_0, p_0)_\omega$ . The surface  $S_0(M)$  is perpendicular to all paths in the set  $\Omega(M)$  of paths in  $M$ .

The invariance of the measure  $d\Gamma$  to Hamiltonian motion and the fact that the velocity field  $\mathbf{v}(q, p)$  in  $\Gamma$  is stationary as the Hamiltonian function  $H(q, p)$  does not depend on time, lead to the following consequence. For any phase space path  $\omega \in \Omega(\Gamma)$ , the product of the velocity  $v((q, p)_\omega)$  and the infinitesimal surface  $dS_\omega$  intersecting the path  $\omega$  perpendicularly at the point  $(q, p)_\omega$ , is constant under Hamiltonian motion along the entire length of the path  $\omega$ , i.e.,

$$v((q, p)_\omega) dS_\omega = \text{const}. \quad (6)$$

For any two points  $(q_0, p_0)_\omega$  and  $(q_a, p_a)_\omega$  on the same path  $\omega$ , the following relation is obtained from (6):

$$v((q_0, p_0)_\omega) dS_{0\omega} = v((q_a, p_a)_\omega) dS_{a\omega}. \quad (7)$$

The infinitesimal element  $dS_{a\omega}$  of the surface  $S_a(M)$  intersects the path  $\omega$  perpendicularly at the point  $(q_a, p_a)_\omega$ . Like the surface  $S_0(M)$ , surface  $S_a(M)$  is also perpendicular to all paths in  $\Omega(M)$ . The infinitesimal elements  $dS_{0\omega}$  and  $dS_{a\omega}$  of the two surfaces  $S_0(M)$  and  $S_a(M)$  are connected by the path  $\omega$  and neighboring paths determined by solutions of Hamilton's equations. The integral over surface  $S_a(M)$  is transformed using (7) into integration over surface  $S_0(M)$ ,

$$\int_{S_a(M)} dS_{a\omega} = \int_{S_0(M)} \frac{v((q_0, p_0)_\omega)}{v((q_a, p_a)_\omega)} dS_{0\omega}. \quad (8)$$

Functional dependence between the points  $(q_0, p_0)_\omega$  and  $(q_a, p_a)_\omega$  on the path  $\omega$  is not explicitly written in the integral (8); it is implied that this functional dependence is determined from solutions of Hamilton's equations.

The following notation is introduced by using (2), with the times  $t$  and  $t_0$  fixed, in the phase function  $f(q, p)$ :

$$f(q(t; q_0, p_0), p(t; q_0, p_0)) \equiv g(q_0, p_0, t_0). \quad (9)$$

Equation (9) is then substituted (with  $t$  and  $t_0$  fixed and the indices in  $(q_0, p_0)$  replaced by the indices  $(q_a, p_a)$ ) into the integral (5), taken over the set  $M$  which is invariant to Hamiltonian motion. This leads to following equality:

$$\int_M f(q, p) d\Gamma = \int_M g(q_a, p_a, t_0) d\Gamma_a. \quad (10)$$

The integral (10) is then transformed using relation (7) and  $d\Gamma_a = ds_{a\omega} dS_{a\omega}$ :

$$\int_M g(q_a, p_a, t_0) ds_{a\omega} dS_{a\omega} = \int_{S_0(M)} dS_{0\omega} v((q_0, p_0)_\omega) \int_\omega \frac{g(q_a, p_a, t_0)}{v(q_a, p_a)} ds_{a\omega}. \quad (11)$$

The function

$$F((q_0, p_0)_\omega, t_0) = v((q_0, p_0)_\omega) \int_\omega \frac{g(q_a, p_a, t_0)}{v(q_a, p_a)} ds_{a\omega}, \quad (12)$$

is defined on the surface  $S_0(M)$  and is called a *path function* or *path distribution*. The integral in the relation (12) defining a path function  $F((q_0, p_0)_\omega, t_0)$  is over the entire length of the path  $\omega$  intersected perpendicularly by the surface  $S_0(M)$  at the point  $(q_0, p_0)_\omega$ . Infinitesimal element of the phase space path  $\omega$  passing through the point  $(q_a, p_a)_\omega$  is  $ds_{a\omega}$ , and the time  $t_0$  in the integral (12) is fixed.

If the phase function  $f(q, p)$  in (10) is a phase space probability density function, equal to zero everywhere outside the invariant set  $M$ , it is straightforward to prove that the path function  $F((q_0, p_0)_\omega, t_0)$  defined by (12) satisfies the nonnegativity and normalization conditions required from probability distributions. Nonnegativity and normalization of the function  $F((q_0, p_0)_\omega, t_0)$  which then represents a *path probability distribution*, follow from the nonnegativity and normalization properties of the related phase space probability density function  $f(q, p)$ . With the help of (10) and (11) and the definition of  $F((q_0, p_0)_\omega, t_0)$  in (12), one then obtains the normalization property

$$\int_M f(q, p) d\Gamma = \int_{S_0(M)} F((q_0, p_0)_\omega, t_0) dS_0 = 1. \quad (13)$$



Nonnegativity is established for all  $(q_0, p_0)_\omega \in S_0(M)$  in a similar way. Integral over any invariant and measurable subset of the set  $M$  is transformed, in the way described above, into integral over a corresponding measurable subset on the surface  $S_0(M)$ . It is clear also that the measure defined on the surface  $S_0(M)$  can be utilized as a measure on the set  $\Omega(M)$  of all phase space paths in some invariant set  $M$ . The correspondence between points  $(q_0, p_0)_\omega \in S_0(M)$  and paths  $\omega \in \Omega(M)$  is one-to-one.

### III. MICROSTATE PROBABILITY AND PATH PROBABILITY

It is now possible to relate the microstate probability and the path probability in the phase space  $\Gamma$  of the system. Let the function  $f(q, p, t)$  be a microstate probability density function on  $\Gamma$ . All points in the phase space  $\Gamma$  move according to Hamilton's equations (1) and  $f(q, p, t)$  satisfies the Liouville equation

$$\frac{\partial f}{\partial t} + \sum_{i=1}^s \left( \frac{\partial f}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \equiv \frac{df}{dt} = 0. \quad (14)$$

Since  $df/dt$  is a total or hydrodynamic derivative, (14) expresses that the time rate of change of  $f(q, p, t)$  is zero along any phase space path given by uniquely determined solution of Hamilton's equations. In the notation used here, this fact is written as

$$f((q, p)_\omega, t) = f((q_0, p_0)_\omega, t_0), \quad (15)$$

where points on the path  $\omega \in \Omega(\Gamma)$  are related by (2).

In addition to the definition of the path probability distribution  $F((q_0, p_0)_\omega, t_0)$  via microstate probability density function  $f(q, p, t)$ , it is possible to give another equivalent definition of  $F((q_0, p_0)_\omega, t_0)$ . In order to accomplish this, probability density function  $\mathcal{F}(q, p, t; q_0, p_0, t_0)$  is introduced on the  $4s$ -dimensional Euclidean space  $\Gamma \times \Gamma$ . This function has the following special properties. If the integral of  $\mathcal{F}(q, p, t; q_0, p_0, t_0)$  is taken over the phase space  $\Gamma$  with  $(q, p)$  as the integration variables, it gives the microstate probability density function  $f(q_0, p_0, t_0)$  at time  $t_0$ ,

$$\int_{\Gamma} \mathcal{F}(q, p, t; q_0, p_0, t_0) d\Gamma = f(q_0, p_0, t_0). \quad (16)$$

Microstate probability density function  $f(q, p, t)$  at time  $t$  is obtained analogously,

$$\int_{\Gamma} \mathcal{F}(q, p, t; q_0, p_0, t_0) d\Gamma_0 = f(q, p, t). \quad (17)$$

It is straightforward to prove, using relation (15), that (16) and (17) are satisfied if the function  $\mathcal{F}(q, p, t; q_0, p_0, t_0)$  has the following form:

$$\mathcal{F}(q, p, t; q_0, p_0, t_0) = f(q, p, t) \prod_{i=1}^s \delta(q_i - q_i(t; q_0, p_0)) \delta(p_i - p_i(t; q_0, p_0)), \quad (18)$$

where  $q_i(t; q_0, p_0)$  and  $p_i(t; q_0, p_0)$  are given by (2) and  $\delta$ -s are Dirac delta functions. In the space  $\Gamma \times \Gamma$  function  $\mathcal{F}(q, p, t; q_0, p_0, t_0)$  given by (18) represents the probability density that the point corresponding to the state of the system is in the element  $d\Gamma_0$  around the point  $(q_0, p_0)$  at time  $t_0$  and in the element  $d\Gamma$  around the point  $(q, p)$  at time  $t$ .

As explained in (17), microstate probability density function  $f(q, p, t)$  is given by the integral of the function  $\mathcal{F}(q, p, t; q_0, p_0, t_0)$  over  $\Gamma$  with  $(q_0, p_0)$  as integration variables. Now, we assume that the set  $M$  of all points in  $\Gamma$  which represent possible microstates of the system is invariant to Hamiltonian motion. By applying the similar procedure and notation that was already introduced in relations (10) and (11), the integral (17) can now be written as

$$f(q, p, t) = \int_{S_0(M)} dS_{0\omega} v((q_0, p_0)_\omega) \int_\omega \frac{\mathcal{F}(q, p, t; q_a, p_a, t_0)}{v(q_a, p_a)} ds_{a\omega}. \quad (19)$$

Along with the lines leading to (19), the function  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  is also introduced:

$$G(q, p, t; (q_0, p_0)_\omega, t_0) = v((q_0, p_0)_\omega) \int_\omega \frac{\mathcal{F}(q, p, t; q_a, p_a, t_0)}{v(q_a, p_a)} ds_{a\omega}. \quad (20)$$

The integral in the definition of  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  in (20) is over the entire length of the phase space path  $\omega$  intersected perpendicularly by the surface  $S_0(M)$  at the point  $(q_0, p_0)_\omega$ . Using (20), relation (19) is then written as

$$f(q, p, t) = \int_{S_0(M)} G(q, p, t; (q_0, p_0)_\omega, t_0) dS_0. \quad (21)$$

It is clear that the expression

$$G(q, p, t; (q_0, p_0)_\omega, t_0) dS_0 d\Gamma \equiv P(q, p, t \cap (q_0, p_0)_\omega, t_0), \quad (22)$$

represents the probability that the point corresponding to the state of the system is at time  $t_0$  anywhere along the paths which pass through an infinitesimal element  $dS_0$  around  $(q_0, p_0)$  on the surface  $S_0(M)$ , and that at some different time  $t$  it is in the volume element  $d\Gamma$  around  $(q, p)$ .

Another definition of the path probability distribution  $F((q_0, p_0)_\omega, t_0)$ , in addition to (12), is now possible in this way. It is given by the integral

$$F((q_0, p_0)_\omega, t_0) = \int_\Gamma G(q, p, t; (q_0, p_0)_\omega, t_0) d\Gamma. \quad (23)$$

Then, in accordance with the theory of probability, the ratio

$$\frac{G(q, p, t; (q_0, p_0)_\omega, t_0) dS_0 d\Gamma}{F((q_0, p_0)_\omega, t_0) dS_0} \equiv P(q, p, t | (q_0, p_0)_\omega, t_0), \quad (24)$$

represents the *conditional probability* that at time  $t$  the point corresponding to the state of the system is in the element  $d\Gamma$  around  $(q, p)$ , if at time  $t_0$  it is anywhere along the paths passing through the infinitesimal element  $dS_0$  around  $(q_0, p_0)$  on the surface  $S_0(M)$ . Relation (23) then proves that the integral of (24) over  $\Gamma$  satisfies the normalization condition, i.e.,

$$\int_{\Gamma} \frac{G(q, p, t; (q_0, p_0)_\omega, t_0) dS_0 d\Gamma}{F((q_0, p_0)_\omega, t_0) dS_0} = 1. \quad (25)$$

To set up all the tools of probability theory needed in this work, conditional probability distribution  $D(q, p, t | (q_0, p_0)_\omega, t_0)$  that corresponds to conditional probability (24), is defined by the relation

$$D(q, p, t | (q_0, p_0)_\omega, t_0) = \frac{G(q, p, t; (q_0, p_0)_\omega, t_0)}{F((q_0, p_0)_\omega, t_0)}. \quad (26)$$

The relation (24), like the relation (22), represents probability which is conserved in the phase space  $\Gamma$ . The total time derivative (i.e., time rate of change along the Hamiltonian flow lines) of this probability is equal to zero. In the relation (24) for the conditional probability, the path probability distribution  $F((q_0, p_0)_\omega, t_0)$  and the surface element  $dS_0$  are independent of the variables  $t$  and  $(q, p)$ . Also, measure  $d\Gamma$  is invariant to Hamiltonian motion. Therefore, it follows that the total time derivative of the conditional probability (24) is equal to zero if and only if

$$\frac{dG}{dt} \equiv \frac{\partial G}{\partial t} + \sum_{i=1}^s \left( \frac{\partial G}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial G}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0. \quad (27)$$

This is a straightforward demonstration that the probability distribution  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  satisfies the equation analogous to the Liouville equation (14) for the microstate probability distribution  $f(q, p, t)$ .

#### IV. INFORMATION ENTROPIES AND MAXENT ALGORITHM

The quantity of the form  $H = - \sum_i p_i \log p_i$  has a central role in information theory as a measure of information, choice and uncertainty for different probability distributions  $p_i$ . In an analogous manner Shannon [4] has defined entropy of a continuous distribution and entropy of  $N$ -dimensional continuous distribution. As pointed out by Jaynes [5], the analog of  $-\sum_{i=1}^n p_i \log p_i$  for a discrete probability distribution  $p_i$  which goes over in a limit of infinite number of points into a continuous

distribution  $w(x)$  (in such a way that the density of points divided by the total number of points approaches a definite function  $m(x)$ ) is given by

$$S_I = - \int w(x) \log \left[ \frac{w(x)}{m(x)} \right] dx. \quad (28)$$

Shannon assumed the analog  $-\int w(x) \log[w(x)]dx$ , but he also pointed out an important difference between his definitions of discrete and continuous entropies. If we change coordinates, the entropy of a continuous distribution will in general change in the way taken into account by Shannon [4]. To achieve the required invariance of entropy of a continuous distribution under a change of the independent variable, it is necessary to introduce the described modification that follows from mathematical deduction [5]. This is achieved with an introduction of the measure function  $m(x)$  and yields the invariant information measure (28). If a uniform measure  $m = \text{const}$  is assumed [5], the invariant information measure (28) differs from the Shannon's definition of entropy of a continuous distribution [4] by an irrelevant additive constant.

Shannon [4] has also defined the joint and conditional entropies of a joint distribution of two continuous variables (which may themselves be multidimensional), concepts that are applied in this work. In the previous section, *joint distribution*  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  of two continuous multidimensional variables  $(q, p) \in \Gamma$  and  $(q_0, p_0)_\omega \in S_0(M)$  was introduced. Following the detailed explanation of (22),  $G(q, p, t; (q_0, p_0)_\omega, t_0)dS_0d\Gamma$  represents the probability of the joint occurrence of two events: the first occurring at time  $t_0$  among the set of all possible phase space paths  $\Omega(M)$  and the second occurring at time  $t$  among the set of all possible phase space points  $M$  which is invariant to Hamiltonian motion. In accordance with Shannon's definition [4], *joint information entropy* of the joint distribution  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  is given by

$$S_I^G(t, t_0) = - \int_{S_0(M)} \int_{\Gamma} G \log G \, d\Gamma dS_0. \quad (29)$$

The notation  $S_I^G(t, t_0)$  indicates that it is a function of times  $t$  and  $t_0$ , through the distribution  $G \equiv G(q, p, t; (q_0, p_0)_\omega, t_0)$ . Following Shannon's definition [4], *conditional information entropy* of the joint distribution  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  is then given by

$$S_I^{DF}(t, t_0) = - \int_{S_0(M)} \int_{\Gamma} G \log \left[ \frac{G}{F} \right] d\Gamma dS_0, \quad (30)$$

where  $F \equiv F((q_0, p_0)_\omega, t_0)$  is the path probability distribution. Using the definition of  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  in (26), one immediately obtains the equivalent form of the conditional information entropy (30):

$$S_I^{DF}(t, t_0) = - \int_{S_0(M)} \int_{\Gamma} DF \log D \, d\Gamma dS_0. \quad (31)$$

From (31) it is clear that the conditional information entropy  $S_I^{DF}(t, t_0)$  is the average of the entropy of conditional probability  $D(q, p, t | (q_0, p_0)_\omega, t_0)$ , weighted over all possible phase space paths  $\omega \in \Omega(M)$  according to the path probability distribution  $F((q_0, p_0)_\omega, t_0)$ .

Relation between the information entropies  $S_I^G(t, t_0)$  and  $S_I^{DF}(t, t_0)$ , introduced in (29) and (30), is completed by introducing the information entropy of the distribution  $F((q_0, p_0)_\omega, t_0)$ , or alternatively, *path information entropy*:

$$S_I^F(t_0) = - \int_{S_0(M)} F \log F \, dS_0. \quad (32)$$

Relation between  $S_I^G(t, t_0)$ ,  $S_I^{DF}(t, t_0)$  and  $S_I^F(t_0)$  is obtained straightforwardly, using (26) in (29), and then applying the properties of probability distributions. In this way one obtains

$$S_I^G(t, t_0) = S_I^{DF}(t, t_0) + S_I^F(t_0). \quad (33)$$

Relation (33), in accordance with the analogous relation of Shannon [4], asserts that the uncertainty (or entropy) of the joint event is equal to the uncertainty of the first plus the uncertainty of the second event when the first is known.

It is important to give some additional comments to (33). In general, uncertainty of the joint event is less than or equal to the sum of uncertainties of the two individual events, with the equality if (and only if) the two events are independent [4]. The probability distribution of the joint event is given here by  $G(q, p, t; (q_0, p_0)_\omega, t_0)$ . Information entropy or uncertainty of one of them (in this case called the second event because of its occurrence at a later time) is equal

$$S_I^f(t) = - \int_\Gamma f \log f \, d\Gamma. \quad (34)$$

The quantity  $S_I^f(t)$  is the information entropy of the microstate probability distribution  $f(q, p, t)$ , or in short, *information entropy*. The uncertainty of the first event is given by the path information entropy  $S_I^F(t_0)$  defined in (32). The aforementioned property of information entropies is given here for  $S_I^G(t, t_0)$ ,  $S_I^f(t)$  and  $S_I^F(t_0)$  by the following relation:

$$S_I^G(t, t_0) \leq S_I^f(t) + S_I^F(t_0), \quad (35)$$

with the equality if (and only if) the two events are independent. Furthermore, from (33) and (35), one obtains an important relation between  $S_I^f(t)$  and  $S_I^{DF}(t, t_0)$ :

$$S_I^f(t) \geq S_I^{DF}(t, t_0), \quad (36)$$

with the equality if (and only if) the two events are independent.

In terms of probability, the events occurring at time  $t_0$  among the set of all possible phase space paths  $\Omega(M)$  and at any time  $t$  among the set of all possible phase space points  $M \subset \Gamma$ , are not independent. If we assume that the values of joint probability distribution  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  are physically well defined (in the sense of (22)) for all points  $(q, p) \in \Gamma$  and  $(q_0, p_0) \in S_0(M)$  at given initial time  $t = t_0$ , its values are then determined at all times  $t$  in the entire phase space  $\Gamma$  via the Liouville equation (27). Simple deduction leads to the conclusion that maximization of the conditional information entropy  $S_I^{DF}(t, t_0)$ , subject to the constraints of Liouville equation (27) and normalization, can not yield the upper bound which is given (at any time  $t$ ) by the value of the information entropy  $S_I^f(t)$  in (36). Attaining this upper bound would require statistical independence or, in other words, a complete loss of correlation between the set of possible phase space paths  $\Omega(M)$  at time  $t_0$  and the set of possible phase space points  $M \subset \Gamma$  at time  $t$ . This is precluded at any time  $t$  by the constraint implied by the Liouville equation (27), and the requirement that the joint probability distribution  $G(q, p, t; (q_0, p_0)_\omega, t_0)$  is well defined.

At this point it is helpful to make a distinction between two aspects of time evolution. The first is a microscopic aspect which represents a problem of dynamics implied in this work by Hamilton's equations. The solutions are represented in  $\Gamma$  as phase space paths. Predicting macroscopic time evolution represents a problem of available information and inferences from that partial information. Therefore, microscopic dynamics and the respective phase space paths are also part of this problem of incomplete information. In a case of macroscopic system, information about microscopic dynamics is very likely to be incomplete for variety of different possible reasons. However, in the absence of more complete knowledge, Hamilton's equations (1) and the set of possible phase space paths are the representation of our information about microscopic dynamics. It is natural to assume that the predicted macroscopic time evolution for a closed system is consistent with our knowledge of microscopic dynamics, even when this knowledge is not complete.

All arguments mentioned before lead to the conclusion that regarding Liouville equation (27) as a strict *microscopic constraint* on time evolution is equivalent to having complete information about microscopic dynamics. Following previously introduced assumptions, the Liouville equation (27) can also be regarded as a *macroscopic constraint* on time evolution. If our information about microscopic dynamics is not sufficient to determine the time evolution, an average is taken over all cases possible on the basis of our partial information. The conditional information entropy  $S_I^{DF}(t, t_0)$  is then maximized subject to the constraint of Liouville equation (27), introduced as a phase space average, or more precisely, an integral over phase space similarly to other macroscopic constraints. In predictive statistical mechanics formulated by Jaynes, inferences are drawn

from probability distributions whose sample spaces represent what is known about the structure of microstates, and maximize information entropy subject to the available macroscopic data [8]. In this way “objectivity” of probability assignments and predictions is ensured from introducing extraneous assumptions not warranted by data. In this work we introduce the same basic idea into stage (2) of the problem of prediction for closed Hamiltonian systems. This approach allows us to consider the incomplete nature of our information about microscopic dynamics in a rational way, and leads to the loss of correlation between the initial phase space paths and final microstates and to uncertainty in prediction. The conditional information entropy  $S_I^{DF}(t, t_0)$  is the measure of this uncertainty, related to loss of information about the state of the system.

## V. MAXENT INFERENCES AND TIME EVOLUTION

In the first approach, time evolution of the conditional probability distribution  $D(q, p, t | (q_0, p_0)_\omega, t_0)$  in the interval  $t_0 \leq t \leq t_a$  should satisfy the following constraints: normalization condition

$$\int_M D(q, p, t | (q_0, p_0)_\omega, t_0) d\Gamma = 1, \quad (37)$$

and the Liouville equation for  $D(q, p, t | (q_0, p_0)_\omega, t_0)$ ,

$$\frac{\partial D}{\partial t} + \sum_{i=1}^s \left( \frac{\partial D}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial D}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0. \quad (38)$$

From (26) it follows that the constraints given by (27) and (38) are equivalent. By definition, the set of all possible microstates  $M \subset \Gamma$  is an invariant set. The normalization constraint (37) contains information about the structure of possible microstates in  $\Gamma$ , in the time interval under consideration  $t_0 \leq t \leq t_a$ . Information about microscopic dynamics is represented by the set  $\Omega(M)$  of possible phase space paths in  $\Gamma$ . In addition, this information is contained in the Liouville equation (38). The assigned path probability distribution  $F((q_0, p_0)_\omega, t_0)$  is compatible with the available information.

Time derivative of the conditional information entropy  $S_I^{DF}(t, t_0)$  in (31) is given by

$$\frac{dS_I^{DF}(t, t_0)}{dt} = - \int_{S_0(M)} \int_M \frac{\partial D}{\partial t} F \log D \, d\Gamma dS_0 - \int_{S_0(M)} \int_M \frac{\partial D}{\partial t} F \, d\Gamma dS_0. \quad (39)$$

Because of the normalization, (37), the last term in (39) is equal to zero. At time  $t_a$ , conditional information entropy  $S_I^{DF}(t_a, t_0)$  is given by the expression,

$$S_I^{DF}(t_a, t_0) = - \int_{t_0}^{t_a} \int_{S_0(M)} \int_M \frac{\partial D}{\partial t} F \log D \, d\Gamma dS_0 dt + S_I^{DF}(t_0, t_0). \quad (40)$$

The following functional is then formed

$$J[D] = S_I^{DF}(t_a, t_0) - S_I^{DF}(t_0, t_0) = \int_{t_0}^{t_a} \int_{S_0(M)} \int_M K(D, \partial_t D) d\Gamma dS_0 dt, \quad (41)$$

with the function  $K(D, \partial_t D)$  given by

$$K(D, \partial_t D) = -\frac{\partial D}{\partial t} F \log D. \quad (42)$$

In the variational problem which is considered here, functional  $J[D]$  in (41) is rendered stationary with respect to variations subject to the constraints (37) and (38). On the boundary of integration region  $M \times (t_0, t_a)$  in the integral (41), function  $D(q, p, t | (q_0, p_0)_\omega, t_0)$  is not required to take on prescribed values. The constraints given by (37) and (38) are written here in equivalent but more suitable form:

$$\varphi_1((q_0, p_0)_\omega, t_0; t, D) = F \int_M D d\Gamma - F = 0, \quad (43)$$

and

$$\varphi_2((q_0, p_0)_\omega, t_0; q, p, t, \partial_q D, \partial_p D, \partial_t D) = \left[ \frac{\partial D}{\partial t} + \sum_{i=1}^s \left( \frac{\partial D}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial D}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \right] F = 0. \quad (44)$$

Methods for variational problems with this type of constraints exist and one can develop them and apply in practical problems [25]. Here, in the notation which is adapted to this particular problem, the following functionals are introduced:

$$C_1[D, \lambda_1] = \int_{S_0(M)} \int_{t_0}^{t_a} \lambda_1 \varphi_1 dt dS_0, \quad (45)$$

and

$$C_2[D, \lambda_2] = \int_{S_0(M)} \int_{t_0}^{t_a} \int_M \lambda_2 \varphi_2 d\Gamma dt dS_0. \quad (46)$$

The Lagrange multipliers  $\lambda_1 \equiv \lambda_1((q_0, p_0)_\omega, t_0; t)$  and  $\lambda_2 \equiv \lambda_2((q_0, p_0)_\omega, t_0; q, p, t)$  are functions defined in the integration regions in (45) and (46). For any function with continuous first partial derivatives, Euler equation for the constraint  $\varphi_2 \equiv \varphi_2((q_0, p_0)_\omega, t_0; q, p, t, \partial_q D, \partial_p D, \partial_t D)$  is equal to zero. Following the multiplier rule for such problems explained in ref. [25], we introduce an additional (constant) Lagrange multiplier  $\lambda_0$  as a multiplicative factor for  $K$ ,

$$J[D, \lambda_0] = \int_{t_0}^{t_a} \int_{S_0(M)} \int_M \lambda_0 K(D, \partial_t D) d\Gamma dS_0 dt. \quad (47)$$

The functional  $I[D, \lambda_0, \lambda_1, \lambda_2]$  is formed from  $J[D, \lambda_0]$ ,  $C_1[D, \lambda_1]$  and  $C_2[D, \lambda_2]$ :

$$I[D, \lambda_0, \lambda_1, \lambda_2] = J[D, \lambda_0] - C_1[D, \lambda_1] - C_2[D, \lambda_2]. \quad (48)$$



Existence of Lagrange multipliers  $\lambda_0 \neq 0$ ,  $\lambda_1$  and  $\lambda_2$ , such that the variation of  $I[D, \lambda_0, \lambda_1, \lambda_2]$  is stationary  $\delta I = 0$ , represents a proof that it is possible to make  $J[D]$  in (41) stationary subject to constraints (43) and (44). The function  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  which renders  $J[D]$  stationary subject to (43) and (44) must satisfy the Euler equation:

$$\lambda_0 \left\{ \frac{\partial K}{\partial D} - \frac{d}{dt} \left( \frac{\partial K}{\partial(\partial_t D)} \right) - \sum_{i=1}^s \left[ \frac{d}{dq_i} \left( \frac{\partial K}{\partial(\partial_{q_i} D)} \right) + \frac{d}{dp_i} \left( \frac{\partial K}{\partial(\partial_{p_i} D)} \right) \right] \right\} - \lambda_1 F + \left[ \frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^s \left( \frac{\partial \lambda_2}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \right] F = 0. \quad (49)$$

It is easy to check that the term multiplied by  $\lambda_0$  in Euler equation (49) is equal to zero. Stationarity of the functional  $I[D, \lambda_0, \lambda_1, \lambda_2]$  in (48) is therefore possible even with  $\lambda_0 \neq 0$ . From (49) it follows that the Lagrange multipliers  $\lambda_1$  and  $\lambda_2$  satisfy the equation

$$\frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^s \left( \frac{\partial \lambda_2}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = \lambda_1. \quad (50)$$

In this variational problem, the function  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  that renders  $J[D]$  in (41) stationary subject to constraints (43) and (44), is not required to take on prescribed values on the boundary of integration region  $M \times (t_0, t_a)$ . Therefore, it is necessary, that in addition to satisfying the Euler equation (49), it also satisfies the Euler boundary condition on the boundary of  $M \times (t_0, t_a)$ , ref. [25]. For all points on the portion of the boundary of  $M \times (t_0, t_a)$  where time  $t = t_0$  or  $t = t_a$ , the Euler boundary condition gives independently:

$$\left[ \frac{\partial K}{\partial(\partial_t D)} - \lambda_2 F \right]_{t=t_0, t_a} = -[\log D + \lambda_2]_{t=t_0, t_a} F = 0. \quad (51)$$

For all points on the portion of the boundary of  $M \times (t_0, t_a)$  where time  $t$  is in the interval  $t_0 < t < t_a$ , the Euler boundary condition gives:

$$F [\lambda_2 \mathbf{v} \cdot \mathbf{n}] \text{ at the boundary of } M = 0. \quad (52)$$

In (52),  $\mathbf{v} \cdot \mathbf{n}$  is a scalar product of the velocity field  $\mathbf{v}(q, p)$  in  $\Gamma$  (defined in Sect. II) and the unit normal  $\mathbf{n}$  of the boundary surface of invariant set  $M$ , taken at the surface. Equation (52) is satisfied naturally due to Hamiltonian motion, since the set  $M$  is invariant by definition, and therefore  $\mathbf{v} \cdot \mathbf{n} = 0$  for all points on the boundary surface of  $M$ . This is a consequence of the fact that phase space paths do not cross over the boundary surface of the invariant set  $M$ .

Functions  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  that render  $J[D]$  in (41) stationary subject to the constraints (43) and (44) are determined from the constraints and the boundary condition given by (51). From

(51) one obtains the form of  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  at times  $t_0$  and  $t_a$ ,

$$D(q, p, t|(q_0, p_0)_\omega, t_0)|_{t=t_0, t_a} = \exp[-\lambda_2((q_0, p_0)_\omega, t_0; q, p, t)]|_{t=t_0, t_a}. \quad (53)$$

Since it is only required that  $t_a \geq t_0$ , time  $t_a$  is arbitrary in other respects. The boundary condition (51) then holds for any time  $t \geq t_0$ :

$$D(q, p, t|(q_0, p_0)_\omega, t_0) = \exp[-\lambda_2((q_0, p_0)_\omega, t_0; q, p, t)]. \quad (54)$$

From the constraint (44), using (54), one immediately obtains an equation for the Lagrange multiplier  $\lambda_2((q_0, p_0)_\omega, t_0; q, p, t)$ :

$$\frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^s \left( \frac{\partial \lambda_2}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0. \quad (55)$$

By comparison of (50) with (55), it follows that for all  $t \geq t_0$ ,

$$\lambda_1((q_0, p_0)_\omega, t_0; t) = 0. \quad (56)$$

As explained in Sect. IV, for any physically well defined conditional probability distribution  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  (in the sense of (24) and (26)), the upper bound on  $S_I^{DF}(t, t_0)$ , given by (36), is not attained in maximization.

The conclusions that follow from the interpretation of (36) and the property of  $S_I^{DF}(t, t_0)$  as a measure of uncertainty explained in Sect. IV, are considered now in the second approach. This is done by replacing the strict equality constraint (44) by the constraint which is of isoperimetric form,

$$\varphi_2((q_0, p_0)_\omega, t_0; t, D) = \int_M \left[ \frac{\partial D}{\partial t} + \sum_{i=1}^s \left( \frac{\partial D}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial D}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \right] F d\Gamma = 0. \quad (57)$$

The functional (46) is then replaced with the functional

$$C_2[D, \lambda_2] = \int_{S_0(M)} \int_{t_0}^{t_a} \lambda_2 \varphi_2 dt dS_0. \quad (58)$$

Lagrange multiplier  $\lambda_2 \equiv \lambda_2((q_0, p_0)_\omega, t_0; t)$  is now a function defined in the integration region in the integral (58). Information that the set  $M$  of possible microstates is invariant to Hamiltonian motion is contained in the constraint (57). Analogy is not complete, because a much larger class of functions satisfies the constraint (57), including all functions that in addition, satisfy also the constraint (44). This fact allows for maximization of the conditional information entropy  $S_I^{DF}(t_a, t_0)$  in (40), subject to constraints (43) and (57), even if  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  is prescribed at initial time  $t_0$ .

The prescribed  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  at initial time  $t_0$  must be physically well defined in the sense of (24) and (26). In this variational problem, function  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  is not required to take on prescribed values on the remaining portion of the boundary of integration region  $M \times (t_0, t_a)$  in (40).

For a function  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  to maximize  $S_I^{DF}(t_a, t_0)$  subject to constraints (43) and (57), it is necessary that it satisfies the Euler equation:

$$\lambda_0 \left\{ \frac{\partial K}{\partial D} - \frac{d}{dt} \left( \frac{\partial K}{\partial(\partial_t D)} \right) - \sum_{i=1}^s \left[ \frac{d}{dq_i} \left( \frac{\partial K}{\partial(\partial_{q_i} D)} \right) + \frac{d}{dp_i} \left( \frac{\partial K}{\partial(\partial_{p_i} D)} \right) \right] \right\} - \lambda_1 F + \frac{\partial \lambda_2}{\partial t} F = 0. \quad (59)$$

Another necessary condition for a maximum, in addition to (59), exists if function  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  is not required to take on prescribed values on a portion of the boundary of  $M \times (t_0, t_a)$ : then, it is necessary that  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  satisfies the Euler boundary condition on the portion of the boundary of  $M \times (t_0, t_a)$  where its values are not prescribed, ref. [25]. In accordance with this, for all points on the portion of the boundary of  $M \times (t_0, t_a)$  where  $t = t_a$ , the Euler boundary condition gives:

$$\left[ \frac{\partial K}{\partial(\partial_t D)} - \lambda_2 F \right]_{t=t_a} = -[\log D + \lambda_2]_{t=t_a} F = 0. \quad (60)$$

The Euler boundary condition is satisfied naturally for all points on the portion of the boundary of  $M \times (t_0, t_a)$  where time  $t$  is in the interval  $t_0 < t < t_a$ . The set  $M$  is invariant to Hamiltonian motion, and equation analogous to (52) is also satisfied here naturally due to Hamiltonian motion.

In analogous manner leading to (50) in the first approach, the Euler equation (59) now leads to the equation for the Lagrange multipliers  $\lambda_1((q_0, p_0)_\omega, t_0; t)$  and  $\lambda_2((q_0, p_0)_\omega, t_0; t)$ :

$$\frac{\partial \lambda_2}{\partial t} = \lambda_1. \quad (61)$$

The form of the MaxEnt conditional probability distribution at time  $t_a$  follows from (60):

$$D(q, p, t_a|(q_0, p_0)_\omega, t_0) = \exp[-\lambda_2((q_0, p_0)_\omega, t_0; t_a)]. \quad (62)$$

For a well defined conditional probability distribution at initial time  $t_0$ , there is an entire class of equally probable solutions  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  obtained by MaxEnt algorithm, which all satisfy the macroscopic constraint (57). At time  $t_a$ , all functions in this class of MaxEnt solutions are equal and given by (62). With the exception of times  $t_0$  and  $t_a$ , the conditional probability distribution  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  obtained by MaxEnt algorithm is not uniquely determined in the interval  $t_0 < t < t_a$ . This is a consequence of the fact that the macroscopic constraint (57) does not determine the time evolution of  $D(q, p, t|(q_0, p_0)_\omega, t_0)$  uniquely, in the way that the strict microscopic

constraint (44) does. However, MaxEnt solutions still predict only time evolutions entirely within the invariant set  $M$ , due to (52). This property follows from the constraint (57), and takes into account the information about the constants of motion that determine the invariant set  $M$ , and in this way, about related conservation laws.

From the normalization (37) of the conditional probability distribution, given at time  $t_a$  by (62), one obtains the relation:

$$W(M) \exp [-\lambda_2((q_0, p_0)_\omega, t_0; t_a)] = 1, \quad (63)$$

where  $W(M)$  is the measure, i.e., phase space volume of the invariant set  $M$ . Equation (63) implies that the Lagrange multiplier  $\lambda_2((q_0, p_0)_\omega, t_0; t)$  at time  $t = t_a$  is independent of the variables  $(q_0, p_0)_\omega$ :

$$\lambda_2((q_0, p_0)_\omega, t_0; t_a) = \lambda_2(t_a). \quad (64)$$

Microstate probability distribution  $f(q, p, t)$  at time  $t = t_a$  is then calculated by using: (21) and (26), the MaxEnt conditional probability distribution  $D(q, p, t | (q_0, p_0)_\omega, t_0)$  at time  $t = t_a$  given by (62) and (64), and the path probability distribution  $F((q_0, p_0)_\omega, t_0)$  at initial time  $t_0$ :

$$f(q, p, t_a) = \exp [-\lambda_2(t_a)]. \quad (65)$$

It follows from (62–65) that at time  $t_a$ , the MaxEnt conditional probability distribution and the corresponding microstate probability distribution are equal,

$$D(q, p, t_a | (q_0, p_0)_\omega, t_0) = f(q, p, t_a) = \exp [-\lambda_2(t_a)] = \frac{1}{W(M)}. \quad (66)$$

From (31), (34) and (66), one obtains the values of information entropies  $S_I^{DF}(t, t_0)$  and  $S_I^f(t)$  at time  $t_a$ ,

$$S_I^f(t_a) = S_I^{DF}(t_a) = \log W(M). \quad (67)$$

Equations (66) and (67) are possible only in case of statistical independence, i.e., the complete loss of correlation between the phase space paths at time  $t_0$ , and the microstates at time  $t_a$ . In general, property of macroscopic systems is that they appear to randomize themselves between observations, provided that the observations follow each other by a time interval longer than a certain characteristic time  $\tau$  called the relaxation time [26]. In the interpretation given here, relaxation time  $\tau$  for a closed Hamiltonian system represents a characteristic time required for the described loss of correlation between the initial phase space paths and final microstates. Furthermore,  $\tau$  also

represents a time interval during which predictions, based on incomplete information about microscopic dynamics, become uncertain to a maximum extent compatible with data. This uncertainty is related to loss of information about the state of the system.

This interpretation is reflected in the role of the Lagrange multipliers  $\lambda_1((q_0, p_0)_\omega, t_0; t)$  and  $\lambda_2((q_0, p_0)_\omega, t_0; t)$ . They are required to satisfy (61), and by integrating it one obtains the following relation,

$$\lambda_2((q_0, p_0)_\omega, t_0; t) = \int_{t_0}^t \lambda_1((q_0, p_0)_\omega, t_0; t') dt' + \lambda_2((q_0, p_0)_\omega, t_0; t_0), \quad (68)$$

for all  $t$  in the interval  $t_0 \leq t \leq t_a$ . By using (68), with (63), (64) and (67), one obtains

$$S_I^f(t_a) = S_I^{DF}(t_a, t_0) = \log W(M) = \int_{t_0}^{t_a} \lambda_1((q_0, p_0)_\omega, t_0; t) dt + \lambda_2((q_0, p_0)_\omega, t_0; t_0). \quad (69)$$

It is clear, from relations (64), (68) and (69), that at time  $t_a$  the Lagrange multiplier  $\lambda_2((q_0, p_0)_\omega, t_0; t_a) \equiv \lambda_2(t_a)$  is determined by the measure  $W(M)$  of the invariant set  $M$  of possible microstates, i.e., the volume of accessible phase space. The subsequent application of MaxEnt algorithm of the described type for a closed system with Hamiltonian dynamics, without the introduction of additional constraints, results in the increase of  $W(M)$ . From (64), (68) and (69) it is then deduced that  $\lambda_2(t_a) \geq \lambda_2(t_0)$ .

Information about the structure of possible microstates restricts the corresponding set, and therefore sets an upper bound on the volume of accessible phase space. The values of  $S_I^{DF}(t_a, t_0)$  and  $S_I^f(t_a)$  at time  $t_a$ , given in (69), are equal to the maximum value of the *Boltzmann-Gibbs entropy*, compatible with this information. The Lagrange multiplier  $\lambda_1((q_0, p_0)_\omega, t_0; t)$ , integrated in (69) over time  $t_0 \leq t \leq t_a$ , is then determined by the rate at which the maximum Boltzmann-Gibbs entropy is attained in reproducible time evolution. The integral in (69), and the quantity  $\lambda_1((q_0, p_0)_\omega, t_0; t)$ , can be identified with the change in entropy, and the *entropy production* for a closed Hamiltonian system, respectively. If information about microscopic dynamics of closed Hamiltonian system is considered complete, whether entropy production can be defined without recourse to coarse graining procedures, or macroscopic, phenomenological approaches, remains an open question. In general, information is discarded in all such models, at some stage, in order to match with what is observed in nature.

## VI. CONCLUSION AND RELATED ISSUES

If we consider the possibility that our information about microscopic dynamics is incomplete, reproducibility and information loss become a part of description of the macroscopic time evolution.

This approach then leads to a simple definition for entropy production. The idea that irreversibility is related to a gradual loss of information has been developed by Jaynes in the density-matrix formalism [2]. Recently, Duncan and Semura [14, 15] suggested the notion that information is really lost at a fundamental level. The interplay of quantum decoherence and dynamics is considered as one of possible reasons behind the second law of thermodynamics, with entropy production caused by information leaking into the environment [27, 28]. In our classical approach, information loss is related to incomplete information about microscopic dynamics. If one considers this idea carefully, even in this simple model, incompleteness of information must be taken into account in some unbiased way.

The issues related to incomplete information can be discussed in an objective manner. Analytical principles for such purposes, i.e. for a separation of the subjective and objective aspects of the theoretical formalism, are found in probability theory. Philosophy of this approach is based on the interpretation of probability theory as a natural extension of deductive logic. Such generalization has been developed in axiomatic way by Jaynes in his treatise on probability theory [29]. It was intended as a tool for plausible inference in situations of incomplete information. The standard axiomatic probability theory is derived from this generalized theory, suggesting in itself that the generalized theory is a proper tool for incorporating new information in our probability distributions. Probability distributions are interpreted in that sense as carriers of incomplete information. This approach is perhaps best understood from descriptions given by Jaynes [29]: "...probability theory as a generalized logic of plausible inference which should apply, in principle, to any situation where we do not have enough information to permit deductive reasoning." We quote also the following lines from [29], which we think are important for the discussion in the next paragraph: "But this is equally true of abstract mathematical systems; when a proposition is undecidable in such a system, that means only that its axioms do not provide enough information to decide it. But new axioms, external to the original set, might supply the missing information and make the proposition decidable after all." We can conclude that when probabilities are interpreted in a related way as a property of our state of knowledge, and applied supplemented with MaxEnt algorithm, that the mathematical description of irreversible behavior fits naturally within the concepts of Shannon's information theory [4].

Another objective aspect of the problem mentioned above is related to the issues that were raised in very interesting and speculative way by Zwick. In his paper [30] on the measurement problem in quantum mechanics, the difficulty of describing it at the level of quantum dynamics is compared, and found to be similar with the incompleteness of certain axiomatic systems in

mathematics, discovered and elaborated by Gödel and others [31, 32]. According to Zwick, the extensive parallelism between the physical and mathematical cases suggest the possibility that the measurement process is self-referential as was Gödel’s special formula, and that measurement may be undecidable within the dynamics (formalism of the time-dependent Schrödinger equation), and occurring only at a meta-level of the formalism. In such line of thinking, physical theory would have then at least two levels; the measurement process would be described at a meta-level, but undecidable on the base level which is described by the dynamical law. At the same time, the base level is inherently incomplete and no contradiction is generated. The dynamical law and all the processes described by it are reversible, but the measurement process is irreversible; in this way irreversibility would be present in a description of the measurement process within such two-level theory. Zwick quotes similar suggestions by Pattee [33] about the necessity of two levels of structure and description for any prediction and control (i.e. measurement) process. Questions that are raised about irreversible behavior of systems governed deterministically by the time-symmetric equations of motion, would then appear paradoxical only in the context of single-level theory [33].

Without involving us more deeply in these issues, we note that in our application of MaxEnt to the problem of prediction of time evolution of closed systems with Hamiltonian dynamics, certain features of two level theory can be clearly recognized. In this simple model they appear only as a result of our recognition of incompleteness of our own information about microscopic dynamics. From pragmatic viewpoint this allows us to discuss further on the related issues about the interplay between our knowledge and measurement constraints on the system and its “actual” dynamics.

## Acknowledgments

Authors wish to thank the anonymous reviewer for insightful suggestions that significantly improved the submitted manuscript. The present work was supported by Croatian MZOS project no. 177-1770495-0476.

- 
- [1] Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. 106, 620–630 (1957)
  - [2] Jaynes, E.T.: Information theory and statistical mechanics. II. Phys. Rev. 108, 171–190 (1957)
  - [3] Gibbs, J.W.: Elementary Principles in Statistical Mechanics. Yale University Press, New Haven (1902)
  - [4] Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423, 623–656 (1948). Reprinted In: Shannon C.E., Weaver, W.: The Mathematical Theory of Communication.

- University of Illinois Press, Urbana (1949)
- [5] Jaynes, E.T.: Information theory and statistical mechanics. In: Ford, K.W. (ed.) 1962 Brandeis Lectures in Theoretical Physics, vol. 3, pp. 181–218. W. A. Benjamin, Inc., New York (1963)
  - [6] Jaynes, E.T.: Gibbs vs Boltzmann entropies. *Am. J. Phys.* 33, 391–398 (1965)
  - [7] Jaynes, E.T.: Where do we stand on maximum entropy? In: Levine, R.D., Tribus, M. (eds.) *The Maximum Entropy Formalism*, pp. 15–118. MIT Press, Cambridge (1979)
  - [8] Jaynes, E.T.: The minimum entropy production principle. *Ann. Rev. Phys. Chem.* 31, 579–601 (1980)
  - [9] Jaynes, E.T.: Macroscopic prediction. In: Haken, H. (ed.) *Complex Systems – Operational Approaches in Neurobiology, Physics, and Computers*, pp. 254–269. Springer, Berlin (1985)
  - [10] Grandy, W.T.: Principle of maximum entropy and irreversible processes. *Phys. Rep.* 62, 175–266 (1980)
  - [11] Garrod, C.: *Statistical Mechanics and Thermodynamics*. Oxford University Press, New York (1995)
  - [12] Jaynes, E.T.: The second law as physical fact and as human inference. Unpublished manuscript (1990). <http://bayes.wustl.edu/etj/node2.html>
  - [13] Zurek, W.H.: Algorithmic randomness and physical entropy. *Phys. Rev. A* 40, 4731–4751 (1989)
  - [14] Duncan, T.L., Semura, J.S.: The deep physics behind the second law: information and energy as independent forms of bookkeeping. *Entropy* 6, 21–29 (2004)
  - [15] Duncan, T.L., Semura, J.S.: Information loss as a foundational principle for the second law of thermodynamics. *Found. Phys.* 37, 1767–1773 (2007)
  - [16] Grandy, W.T.: Time evolution in macroscopic systems. I. Equations of motion. *Found. Phys.* 34, 1–20 (2004)
  - [17] Grandy, W.T.: Time evolution in macroscopic systems. II. The entropy. *Found. Phys.* 34, 21–57 (2004)
  - [18] Grandy, W.T.: Time evolution in macroscopic systems. III. Selected applications. *Found. Phys.* 34, 771–813 (2004)
  - [19] Tishby, N.Z., Levine, R.D.: Time evolution via a self-consistent maximal-entropy propagation: the reversible case. *Phys. Rev. A* 30, 1477–1490 (1984)
  - [20] Plastino, A.R., Plastino, A.: Statistical treatment of autonomous systems with divergenceless flows. *Physica A* 232, 458–476 (1996)
  - [21] Plastino, A., Plastino, A.R., Miller, H.G.: Continuity equations, H-theorems, and maximum entropy. *Phys. Lett. A* 232, 349–355 (1997)
  - [22] Plastino, A.R., Plastino, A.: Universality of Jaynes’ approach to the evolution of time-dependent probability distributions. *Physica A* 258, 429–445 (1998)
  - [23] Schönfeldt, J.-H., Jimenez, N., Plastino, A.R., Plastino, A., Casas, M.: Maximum entropy principle and classical evolution equations with source terms. *Physica A* 374, 573–584 (2007)
  - [24] Khinchin, A.I.: *Mathematical Foundations of Statistical Mechanics*. Dover, New York (1949)
  - [25] Wan, F.Y.M.: *Introduction to the Calculus of Variations and Its Applications*. Chapman & Hall, New York (1995)
  - [26] Kittel, C.: *Elementary Statistical Physics*. Wiley, New York (1958)



- [27] Zurek, W.H., Paz, J.-P.: Decoherence, chaos, and the second law. *Phys. Rev. Lett.* 72, 2508–2512 (1994)
- [28] Zurek, W.H.: Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.* 75, 715–775 (2003)
- [29] Jaynes, E.T.: *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge (2003). Bretthorst, G.L. (ed.)
- [30] Zwick, M.: Quantum measurement and Gödel’s proof. *Speculat. Sci. Technol.* 1, 135–145 (1978)
- [31] Gödel, K.: *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. Basic Books, Inc., New York (1962)
- [32] Nagel, E., Newman, J.R.: *Gödel’s Proof*. New York University Press, New York (1960)
- [33] Pattee, H.H.: Postscript. In: Pattee, H.H. (ed.) *Hierarchy Theory*, pp. 131–156. George Braziller, New York (1973)